Design of the Intel Pentium™ Processor

Avtar Saini

Intel Corp.2200 Mission College Blvd.,
RN2-27, Santa Clara, Ca. 95052

1.0 Abstract

The Intel Pentium processor is the next generation high performance processor, fully compatible with it's predecessor, the i486™. The Pentium processor integrates 3.1Million transistors on 0.8u BiCMOS process technology and is 2-5X higher performance than the i486 CPU at the same frequency. Techniques used to achieve this level of performance are described.

2.0 Micro-architecture overview

The design of the Pentium processor started in early 1989 with the primary goal of maximizing performance while preserving software compatibility within the practical constraints of available technology. As such, the Pentium processor is fully binary compatible with the i486. This allows the large existing software base of the i486 to execute, without any modifications on the Pentium processor, albeit faster. The speed up is the result of a totally new Superscaler micro-architecture design.

The core execution units are two integer pipelines and a floating point pipeline with dedicated adder, multiplier and divider (see figure 2 on the next page). To help feed the fast decode and execution engines, the Pentium processor has separate code and data caches, with a branch target buffer augmenting the instruction cache for dynamic branch prediction. The external interface includes separate address and 64 bit data bus.

The Pentium processor integer pipeline has the same basic five stage structure (see figure 1) introduced on the i486 processor i.e. (i) Prefetch and align - PF, (ii) First decode - D1, (iii) Second decode and address generation - D2, (iv) Execute -E and, (v) Write back - WB.

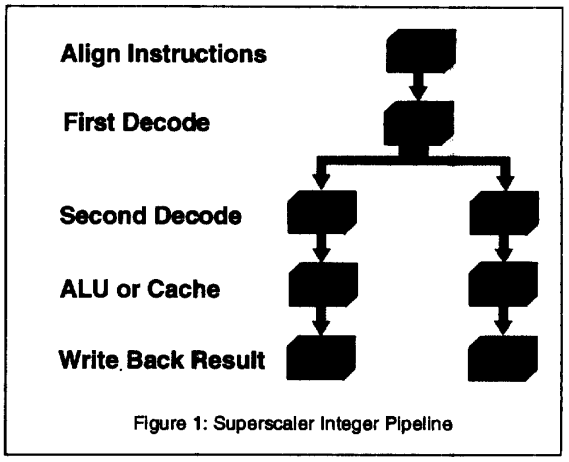

Figure 1: Superscaler Integer Pipeline

1. Dual instruction decode and execution: The Pentium processor can decode two consecutive "Simple" instructions I1 and I2.

If the following are all true,
I1 is a "Simple" instruction
I2 is a "Simple" instruction
I1 is not a JUMP instruction
Destination of I1 is not the Source of I2
Destination of I1 is not the Destination of I2
Then both I1 and I2 are issued to the U-pipe and the V-pipe
Else only I1 is issued to the U-pipe.
"Simple" instruction are generally ALU or MOV operations, including Reg-Reg, Imm-Reg, Mem-Reg, and Reg-

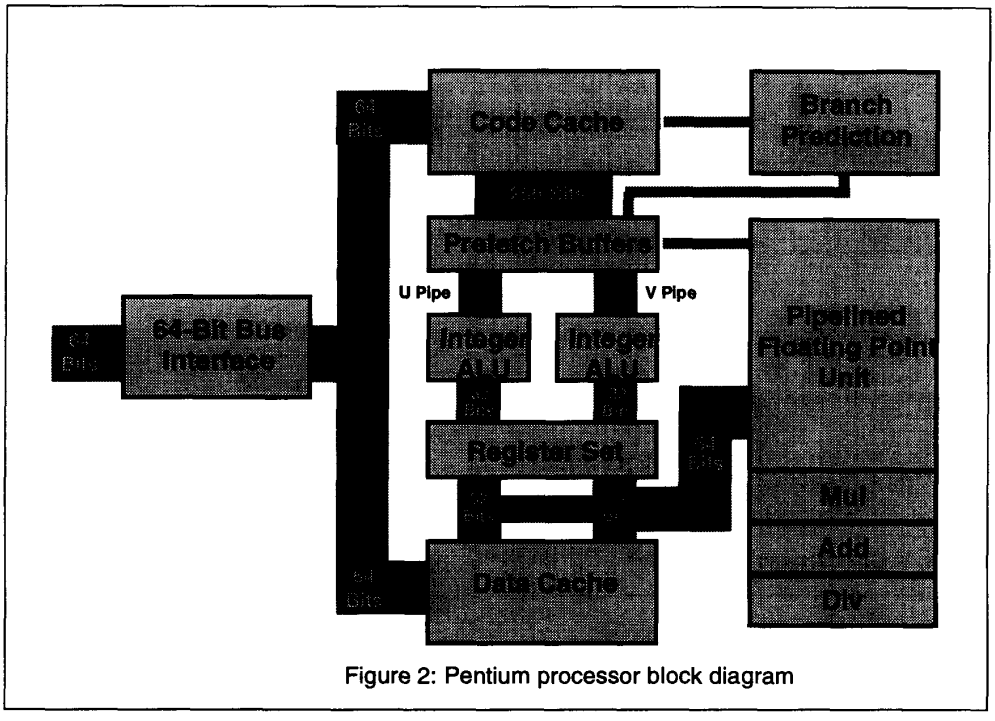

Figure 2: Pentium processor block diagram

Mem Formats, and JUMPS. More than 90% of the instructions executed in the Integer SPEC benchmark suite are simple.

2. Superscaler execution: The integer datapath is an integrated implementation of two pipelines, U and V with independent access to the data cache, independent address generation and two complete ALUs. In the PF and D1 stages the processor can fetch and decode two simple instructions in parallel and issue them to the U and V pipelines. Additionally, for complex instructions the CPU in D1 can generate microcode sequences that control both the U and V pipelines.

3. Separate code and data caches: To help feed the fast decode and execution engines the Pentium processor has separate code and data cache, which are each 8K-byte, two-way associative designs with 32 byte line size. Separate caches help eliminate prefetcher versus data access conflicts. And, efficient branch prediction requires that the destination of a branch be accessed simultaneously with data references of previous instructions executing in the pipeline. Separate caches do require extra logic to handle self-modifying code, however.

Programs executing on the i486 CPU typically generate more data memory references than when executing on RISC processor. Measurements on Integer SPEC benchmarks show 0.5 to 0.6 data references per instruction for the i486 and only 0.17 to 0.33 for the MIPS processor. This difference results directly from the limited number (eight) of registers for the x86 architecture, as well as procedure calling conventions that require passing all parameters in memory. A small data cache is adequate to capture the locality of the additional references. (After all, the additional references have sufficient locality to fit in the register file of the RISC microprocessors). The Pentium processor implements a data cache that supports dual accesses by the U and V pipes to provide additional bandwidth and simplify compiler instruction scheduling algorithms.

As shown in figure 3,The data cache array itself is implemented as a single ported 8-way interleaved memory. The TLB and the cache tags are dual ported. The U and the V ports are 32-bit each and are combined to form a unified 64-bit floating point port.

4. Branch prediction: The Pentium processor employs a branch target buffer (BTB), which is an associative memo-

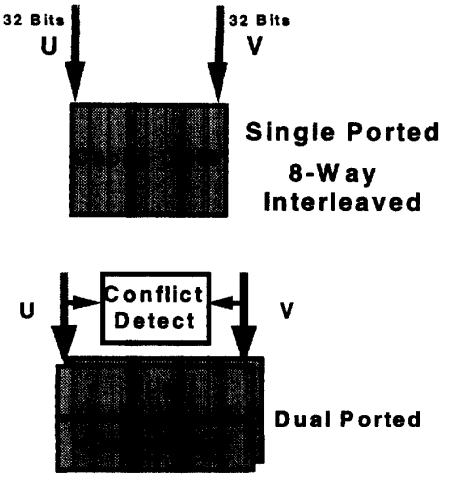

Figure 3: Dual access data cache

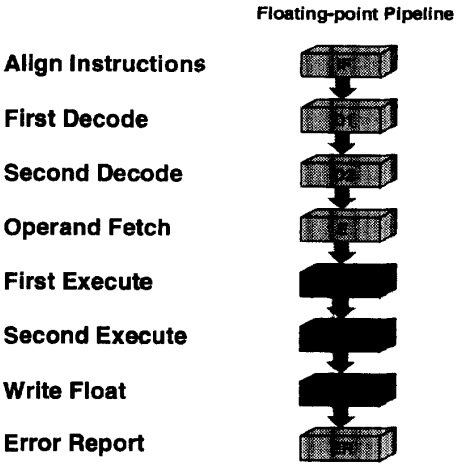

Figure 4: Floating-point Pipeline

ry used to improve performance of taken branch instructions. When branch is first taken, the CPU allocates an entry in the branch target buffer to associate the branch instruction's address with its destination address and to initialize the history used in the prediction algorithm. As instructions are decoded. the processor searches the BTB to determine whether it holds and entry for a corresponding branch instruction. When there is a hit, the processor uses the history to determine whether the branch should be taken. If it is, the processor uses the target address to begin fetching and decoding instructions from the target path. The branch is resolved early in the WB stage, if the prediction was incorrect, the CPU flushes the pipeline and resumes fetching from the correct path. The branch target buffer holds entries for 256 branches in a four way associative organization. Correctly predicted branches execute with no delays. Branch prediction is software transparent and benefits all existing code.

5. **High performance floating point:** The floating point pipeline (see figure 4) is eight stages, where the first five are shared with the integer pipeline. Prefetch (PF), First decode (D1), Second decode and address generation (D2), Operand fetch and memory write if FP store instruction(E), first execute (X1), second execute (X2), rounding and write result to the FP register file (WF), and Error reporting (ER).

The eight stage pipeline allows a single cycle throughput for most of the "basic" floating point instructions such as add, subtract, multiply and, compare. Floating point instructions execute in the U pipe and generally cannot be paired with any other integer or floating point instructions.

3.0 Data Integrity and Functional Redundancy

In addition to the bus, all major arrays on chip i.e. the code and data caches, cache tags, TLBs, microcode ROM and branch target buffer are covered by internal parity.

Functional redundancy check (FRC) is implemented. Two Pentium processors can be used as Master and Checker to increase system reliability and enable new applications.

4.0 New Compiler Technology

To help achieve the most optimum performance, new compiler technology was developed along side the Pentium processor. It incorporates well known machine independent optimizations common to current high performance compilers, such as inlining, unrolling, and loop transformations. In addition, x86 specific techniques were developed and tuned for the Pentium processor micro-architecture.

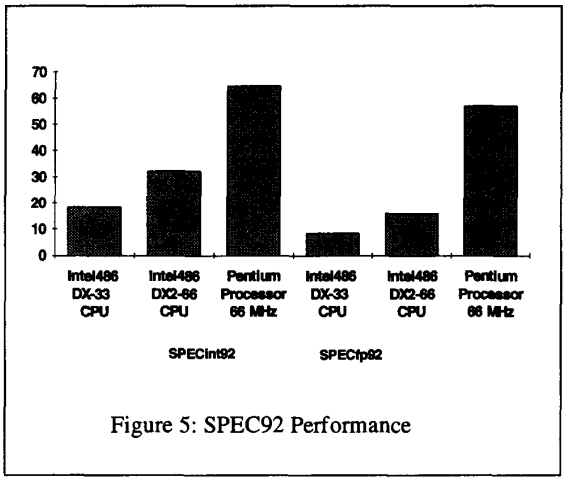

Figure 5: SPEC92 Performance

The x86 architecture supports a variety of instruction formats for equivalent operations. Consequently, it is critical to select instruction formats that are decoded most efficiently by the processor. The x86 register set includes only eight registers, common global register allocation techniques that assign variables to registers for the entire scope of a procedure are ineffective with such a limited number of registers. Registers must be allocated within a narrower scope and together with instruction scheduling. The compiler schedules instructions to minimize interlocks and to maximize parallel execution for the superscaler pipelines.

5.0 Validation

One of major challenges in the making of the Pentium processor was its validation. Major advancements were made both in pre-silicon and post-silicon testing. Pre-silicon logic validation has four focus areas.

1. Architecture verification or "black box" testing, functionality is checked from the programmer's point of view. The Pentium processor was tested using the extensive test suites developed over time at Intel which cover all aspects of the programming model and all the Pentium processor user visible features.

2. Design verification or "white box" testing, where the validation team indulged in the logic details of the micro-architecture and tested various components e.g. state machines and inter-unit protocols extensively. The idea here was to extensively test smaller components of the design.

3. Random instruction testing: The Pentium logic model was compared to a software emulator using vectors from a random code generator. Random instruction testing was a valuable tool to cover all those situations that are rarely covered by the more traditional, handwritten tests. The complete architecture space was exercised extensively using this means.

4. QuickTurn: The Pentium processor model was ported onto a quickturn setup (a logic design hardware model) comprising of an array of rapid prototyping modules (RPMs) capable of handling the complete model. The memory arrays were built using discrete circuits. QuickTurn allowed validation at a significantly higher speed than was possible with the software model. QuickTurn helped validate major operating systems and applications before committing silicon.

Post silicon testing comprised of silicon characterization on the production tester, system validation using engineering systems designed to exercise various bus configurations, extensive testing by the alpha and beta sites and compatibility testing.

6.0 Summary

The Pentium Processor employs superscaler integer pipeline, branch prediction, and highly pipelined floating point unit to achieve the highest x86 performance while preserving binary compatibility with the x86 architecture. Figure 5 summarizes the performance of the Pentium processor and the highest performance i486 CPU for the SPEC benchmarks in well-tuned systems.

7.0 References

[1] Pentium Processor User's Manual, *Volume 1: Pentium Processor data book, Volume 3: Architecture and Programming model,* Intel Corporation, Santa Clara, California, 1993